

Remarks

Claims 1-21 are pending in the application.

In many instances, an organization receiving data records from multiple data providers finds that each data provider formats its data records differently. This can create a significant challenge for the organization which must both reformat each record received from the various sources and identify and eliminate repetitive data records. For example, an entity such as an information assistance service provider, which receives from various telecommunications companies (e.g., AT&T, SBC, etc.) data records containing customers' addresses, telephone numbers, etc., typically must incorporate all of the data received into its database(s) while ensuring that repetitive data is eliminated. When the data records received from the telecommunication companies are in different formats (as is typical), the task of combining the data only becomes greater.

The invention is directed to a system and method for incorporating data received from different sources into one or more databases having records in a specified format. Data (e.g., records containing names and telephone numbers) are received from multiple sources (p. 6, lines 1-3). In accordance with the invention, a converter routine converts each received record into a data record conforming to a uniform format (p. 6, lines 7-10). For example, the uniform format may require, say, that each record comprise a first field containing an identifier, a second field containing a telephone number, a third field containing an address, etc. (Figs. 4C-4D). A normalizer then modifies the contents of the fields within each record, as necessary, to conform to a predetermined nomenclature. Thus, data in the address field of one record containing the letters "California" may be changed to "CA," while data in another record containing the letters "Ca" may also be changed to "CA" (p. 9, lines 15-20).

After the received data records are reformatted and normalized, a processor identifies sets of potentially equivalent normalized data records (p.14, lines 11-15). For example, two normalized data records containing identical telephone numbers (in their respective telephone number fields) may be identified as records that may represent a single listing (p.14, lines 11-15). The processor then determines whether the identified records are actually equivalent (p. 14, lines 24-25). This is achieved by comparing each of the two normalized data records to the other

Serial No. 09/992,987

normalized data record on a field-by-field basis (p. 15, lines 5-16). For example, the processor may compare the two records on a field-by-field basis and calculate a confidence value reflecting how many fields are identical (p. 15, lines 5-16). If the confidence level value is sufficiently high, the data records are deemed to be equivalent (p.15, lines 12-16).

In accordance with an aspect of the invention, when two records are deemed equivalent, the processor may create a final record in the uniform format (p. 16, line 16 - p. 17, line 6). For each defined field within the final record, data items from corresponding fields of the equivalent normalized records are grouped together and compared, and data items from one of the records is selected (p. 16, line 14 - p. 17, line 6). Comparisons between data items in corresponding fields may be conducted based on, e.g., reliability rankings (p. 16, lines 14-26). The selected data item is copied to the corresponding field in the final record.

Claim Rejections - 35 USC § 103

Claims 1-21 were rejected under 35 U.S.C. 103(a) as being allegedly unpatentable over Kane in view of Seilhamer. In response, applicant has amended claims 1, 3, 4, 7, 10, 11, 14, 15, 17, 18, and 21.

Kane discloses a system and method for utilizing data stored in one or more source databases to update data stored in a target database. Data records in the source databases are converted to a common format (col. 7, lines 13-20). The reformatted source records are then sequentially processed to update the target database (col. 9, lines 30-32). Specifically, a source record is selected, and a search is performed to identify a matching record in the target database (col. 9, lines 60-65). Matching is performed using a "sliding window" method that compares text in two records on a character-by-character basis (col. 11, lines 1-8). If a match is found, corresponding data in the two records are ranked according to a confidence measure (col. 8, lines 25-40). If the data in the source record has a higher rank than the corresponding data in the target record, the data from the source record is copied to the target record (col. 12, lines 40-45).

Seilhamer discloses a relational database system for storing biomolecular sequence information in a manner that allows sequences to be catalogued and searched according to one or more protein function hierarchies (col. 2, lines 15-20). The hierarchies are provided to allow

carefully tailored searches for sequences based upon a protein's biological or molecular function (col. 2, lines 15-22).

Amended Claim 1

Nowhere do Kane and Seilhamer, individually or in combination, teach or suggest selecting data items from (a) a first assemblage or (b) a second assemblage when a "particular" data item (e.g., a telephone number) in the first assemblage is "identical" to a corresponding data item in the second assemblage, as amended claim 1 now recites. In fact, Kane teaches away from the invention by using a "sliding window" technique comparing every character in each word in a source record with every character in each word of a target record (col. 11, lines 1-5), and by determining whether or not to copy data only after all the characters and words in both records have been examined. By contrast, in accordance with the invention represented by amended claim 1, the determination of the above selection of data items from (a) or (b) is made as soon as the particular data item in the first assemblage and the corresponding data item in the second assemblages are found identical.

Seilhamer also fails to teach or suggest these features. As such, amended claim 1, together with its dependent claims (2-3) is patentable over the cited art.

Amended claim 4 and amended claim 1 share similar claim elements. Thus, for reasons set forth above for amended claim 1, amended claim 4, together with its dependent claims (5-6), is patentable over the cited art.

Amended Claim 7

In addition, Kane and Seilhamer, individually and in combination, fail to disclose or suggest "determining a [score] value representing a number of corresponding fields in the first assemblage and the second assemblage having identical data items," as amended claim 7 now recites. Although Kane discloses "using two sliding windows that move along each word trying to mate up characters and patterns" (col. 11, lines 1-2), "scoring of compares is done by starting with a score equal to the sum of the length of the two words and subtracting points for each character that is not the same" (col. 11, lines 2-5). Thus, Kane discloses determining a score

based on the number of identical characters in words, rather than identical data items in fields as amended claim 7 now recites. That is, for example, in the claimed invention a score “0” would be registered if some, but not all, characters in the corresponding data items are identical, while in Kane, in that case the score would be non-zero because of some identical characters.

A fortiori, Kane fails to teach or suggest a “data masher based on the [score] value selecting at least one of the data items in the first assemblage and the second assemblage to form the record,” as amended claim 7 also recites. Seilhamer likewise fails to teach or suggest these features. As such, amended claim 7, together with its dependent claims (8-10), is patentable over the cited art.

Amended Claim 11

Amended claim 11, like amended claim 7, recites a score “value” representing a number of corresponding fields in a first record and a second record having “identical data items” therein. As discussed above, Kane and Seilhamer fail to teach or suggest this feature. Accordingly, amended claim 11, together with its dependent claims (12-14), is patentable over the cited art.

Amended Claims 15-21

Amended claim 15 and amended claim 1 share similar claim elements. Thus, for reasons set forth above for amended claim 1, amended claim 15, together with its dependent claims (16-17), is patentable over the cited art.

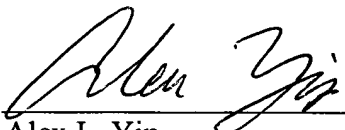
Amended claim 18 and amended claim 11 share similar claim elements. Thus, for reasons set forth above for amended claim 11, amended claim 18, together with its dependent claims (19-21), is patentable over the cited art.

Serial No. 09/992,987

Conclusion

In view of the foregoing, each of claims 1-21, as amended, is believed to be in condition for allowance. Accordingly, reconsideration of these claims is requested and allowance of the application is earnestly solicited.

Respectfully,

By 
Alex L. Yip
Attorney for Applicant
Reg. No. 34,759
212-836-7363

Date: June 23, 2004